

ADMM アルゴリズムを用いた ILRMA 型ブラインド音源分離*

© 渡會博子 (農工大), 松本和樹 (早大), 山田宏樹, 矢田部浩平 (農工大)

1 まえがき

優決定ブラインド音源分離 (BSS) では, 音源の特徴を考慮したさまざまな音源モデルを設計することで分離性能の向上が図られてきた. 独立ベクトル分析 (IVA)[1] や独立低ランク行列分析 (ILRMA)[2] は, それぞれ音源のグループスパース性および振幅スペクトログラムの低ランク性を考慮した音源モデルを仮定し, 高い分離性能を実現している. より現実に忠実な音源モデルを設計できれば, さらに高性能な音源分離手法が得られる可能性がある.

我々はこれまで, 音源モデル設計の自由度を高めるため, 交互方向乗数法 (ADMM) を用いた優決定 BSS フレームワークを提案してきた [3]. ADMM アルゴリズムは非凸最適化問題に対して有効な最適化アルゴリズムの一つである. また, 既存の問題に対して正則化項や制約を追加する際も, 変数分離によって容易に対応できる. このフレームワークを用いて実現した IVA (ADMM-IVA) は, AuxIVA と同等またはそれ以上の分離性能を示すことが確認されている [3].

本稿では, このフレームワークを拡張し, ADMM を用いた ILRMA 型 BSS フレームワークを提案する. 提案手法は, 従来の ILRMA と同様に分離音の振幅スペクトログラムの非負値行列因子分解 (NMF) による近似を導入することで, 分離音の低ランク性を考慮した音源モデル設計を可能とする. また, 初期検討として, 分離音の振幅とその NMF による低ランク近似の構造を関連付ける正則化項に perspective 関数を用いる. 本稿で用いる perspective 関数は凸関数であり, 近接作用素が解析的に求められるため [4], アルゴリズムの導出が容易である. また, 凸な正則化項を用いることにより, 初期値依存性の低減が期待できる. 実験では, perspective 関数を用いた ILRMA 型 BSS と従来の ILRMA との性能的な違いを調査した. 結果として, 提案手法は適切なパラメータ設定の下で従来手法と同程度の分離性能を達成し, また分離に失敗するケースを減らせることが確認された.

2 従来手法

2.1 ADMM アルゴリズムを用いた優決定 BSS

多チャンネル BSS は, 分離行列 $\mathbf{W}[f] \in \mathbb{C}^{N \times M}$ の推定問題として定式化可能である. ここで, N は音源数, M はマイクロホン数である. マイクロホン数 M が音源

Algorithm 1 ADMM-BSS

Input: $\mathbf{X}, \mathbf{w}^{[1]}, \mathbf{z}_1^{[1]}, \mathbf{z}_2^{[1]}, \mathbf{u}_1^{[1]}, \mathbf{u}_2^{[1]}, \rho, \alpha$
Output: $\mathbf{w}^{[K+1]}$

- 1: **for** $k = 1, \dots, K$ **do**
- 2: $\mathbf{w}^{[k+1]} = (\mathbf{X}^H \mathbf{X} + \mathbf{I})^{-1} ((\mathbf{z}_1^{[k]} - \mathbf{u}_1^{[k]}) + \mathbf{X}^H (\mathbf{z}_2^{[k]} - \mathbf{u}_2^{[k]}))$
- 3: $\boldsymbol{\xi}_1 = \alpha \mathbf{w}^{[k+1]} + (1 - \alpha) \mathbf{z}_1^{[k]}$
- 4: $\boldsymbol{\xi}_2 = \alpha \mathbf{X} \mathbf{w}^{[k+1]} + (1 - \alpha) \mathbf{z}_2^{[k]}$
- 5: $\mathbf{z}_1^{[k+1]} = \text{prox}_{\frac{\rho}{\alpha}} [\boldsymbol{\xi}_1 + \mathbf{u}_1^{[k]}]$
- 6: $\mathbf{z}_2^{[k+1]} = \text{prox}_{\frac{\rho}{\alpha}} [\boldsymbol{\xi}_2 + \mathbf{u}_2^{[k]}]$
- 7: $\mathbf{u}_1^{[k+1]} = \mathbf{u}_1^{[k]} + \boldsymbol{\xi}_1 - \mathbf{z}_1^{[k+1]}$
- 8: $\mathbf{u}_2^{[k+1]} = \mathbf{u}_2^{[k]} + \boldsymbol{\xi}_2 - \mathbf{z}_2^{[k+1]}$
- 9: **end for**

数 N 以上である優決定 BSS では, 適切な分離行列が推定できれば, 時間周波数領域における M チャンネルの観測信号 $\mathbf{x}[t, f] = [x_1[t, f], \dots, x_M[t, f]]^T \in \mathbb{C}^M$ から, N チャンネルの分離音 $\mathbf{y}[t, f] = [y_1[t, f], \dots, y_N[t, f]]^T \in \mathbb{C}^N$ を $\mathbf{y}[t, f] = \mathbf{W}[f] \mathbf{x}[t, f]$ によって得られる. 従来の優決定 BSS 手法の多くは, 以下の最適化問題の解として分離行列を得る [5].

$$\underset{(\mathbf{W}[f])_{f=1}^F}{\text{Minimize}} \mathcal{P}(\mathbf{W}[f] \mathbf{x}[t, f]) - \sum_{f=1}^F \log |\det(\mathbf{W}[f])| \quad (1)$$

ただし, F は周波数ビン数であり, $\mathcal{P} : \mathbb{C}^{NM} \rightarrow \mathbb{R} \cup \{+\infty\}$ は音源モデルに対応する関数である. \mathcal{P} が ℓ_1 ノルムの場合は独立成分分析 (ICA) に, $\ell_{2,1}$ 混合ノルムの場合は IVA に対応する. 式 (1) を ADMM によって最適化可能な形式に再定式化すると,

$$\begin{aligned} & \underset{(\mathbf{w}, \mathbf{z}_1, \mathbf{z}_2)}{\text{Minimize}} \quad \mathcal{I}(\mathbf{z}_1) + \mathcal{P}(\mathbf{z}_2) \\ & \text{subject to} \quad \mathbf{z}_1 = \mathbf{w}, \quad \mathbf{z}_2 = \mathbf{X} \mathbf{w} \end{aligned} \quad (2)$$

となる. ここで, $\mathbf{w} \in \mathbb{C}^{NM}$ は分離行列 $\mathbf{W}[f]$ を $\mathbf{w}[f] = \text{vec}(\mathbf{W}[f]) = [W_{1,1}[f], \dots, W_{1,M}[f], W_{2,1}[f], \dots, W_{N,M}[f]]^T \in \mathbb{C}^{NM}$ のようにベクトル化し, それらを全周波数について連結したものである. また, $\mathbf{X} \in \mathbb{C}^{NTF \times NMF}$ は観測信号からなるブロック対角行列であり, \mathcal{I} は $\mathcal{I}(\mathbf{w}) = -\sum_{f=1}^F \sum_{n=1}^N \log \sigma_n(\text{mat}(\mathbf{w}[f]))$ である. ただし, $\sigma_n(\cdot)$ は行列の n 番目の特異値を得る関数, $\text{mat}(\cdot)$ は $\text{vec}(\cdot)$ の逆作用素である. また, $\mathbf{z}_1, \mathbf{z}_2$ は補助変数である. Alg. 1 は式 (2) に ADMM アルゴリズムを適用したものである. ただし, $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$ の更新式はアルゴリズムの安定性に関わる緩和操作であり, $\alpha \in (0, 2)$, $\rho > 0$ である. また, $\mathbf{u}_1, \mathbf{u}_2$ は双対変数であり, $(\cdot)^H$ は複素共役転置である [3].

*ILRMA-like blind source separation using ADMM algorithm. By Hiroko WATARAI (Tokyo University of Agriculture and Technology), Kazuki MATSUMOTO (Waseda University), Koki YAMADA and Kohei YATABE (Tokyo University of Agriculture and Technology).

2.2 独立低ランク行列分析

ILRMA は音源の振幅スペクトログラムの低ランク性を考慮した音源モデルを仮定しており、分離音の振幅に対して NMF による低ランク近似を取り入れる [2]。ILRMA は、以下の問題として定式化できる。

$$\begin{aligned} & \underset{(\mathbf{w}, \mathbf{T}, \mathbf{V})}{\text{Minimize}} && \mathcal{I}(\mathbf{w}) + \mathcal{D}(\mathbf{X}\mathbf{w}, \mathbf{T}\mathbf{V}) \\ & \text{subject to} && \mathbf{T} \geq 0, \quad \mathbf{V} \geq 0 \end{aligned} \quad (3)$$

ただし、 $\mathbf{T} = (\mathbf{T}_n)_{n=1}^N$, $\mathbf{V} = (\mathbf{V}_n)_{n=1}^N$ は、基底行列 $\mathbf{T}_n \in \mathbb{R}_+^{F \times R}$ およびアクティベーション行列 $\mathbf{V}_n \in \mathbb{R}_+^{R \times T}$ の組であり、行列の各要素に対し非負制約が課されている。ここで、 R は NMF の基底数である。また、チャンネルごとの行列積 $(\mathbf{T}_n \mathbf{V}_n)_{n=1}^N$ によって得られる低ランク近似を $\mathbf{T}\mathbf{V}$ と表す。関数 \mathcal{D} には、分離音 $\mathbf{X}\mathbf{w}$ の振幅スペクトログラムと低ランク近似 $\mathbf{T}\mathbf{V}$ の間に距離やそれに類似した尺度を与える関数を用いる。従来の ILRMA では、関数 \mathcal{D} として局所ガウス分布の負対数尤度関数に対応する非凸関数が用いられる。この関数は、NMF 変数 \mathbf{T}, \mathbf{V} に関しては定数項の違いを除いて ISNMF のコスト関数と一致する [2]。

3 提案手法

本稿では、従来手法 [3] の拡張として、ADMM を用いた ILRMA 型 BSS フレームワークを提案する。提案手法は、正則化項や制約の追加が容易であり、数値的に不安定な逆行列演算や除算を回避できる。また本稿では、関数 \mathcal{D} として近接作用素が解析的に求められる perspective 関数を利用する [4]。

3.1 ILRMA 型 BSS フレームワーク

式 (3) は、5 つの補助変数 $\tilde{\mathbf{w}} \in \mathbb{C}^{NMF}$, $\mathbf{Y} \in \mathbb{C}^{NTF}$, $\mathbf{\Upsilon} \in \mathbb{R}^{NTF}$, $\tilde{\mathbf{T}} \in \mathbb{R}^{NFR}$, $\tilde{\mathbf{V}} \in \mathbb{R}^{NRT}$ および \mathcal{D} の重み λ を導入し変数分離することで [6]、次式の通りになる。

$$\begin{aligned} & \underset{\tilde{\mathbf{w}}, \mathbf{Y}, \mathbf{\Upsilon}, \tilde{\mathbf{T}}, \tilde{\mathbf{V}}}{\text{Minimize}} && \mathcal{I}(\tilde{\mathbf{w}}) + \lambda \mathcal{D}(\mathbf{Y}, \mathbf{\Upsilon}) \\ & \text{subject to} && \tilde{\mathbf{w}} = \mathbf{w}, \quad \mathbf{Y} = \mathbf{X}\mathbf{w}, \quad \mathbf{\Upsilon} = \mathbf{T}\mathbf{V}, \\ & && \mathbf{T} = \tilde{\mathbf{T}}, \quad \mathbf{V} = \tilde{\mathbf{V}}, \quad \tilde{\mathbf{T}} \geq 0, \quad \tilde{\mathbf{V}} \geq 0 \end{aligned} \quad (4)$$

ここで、 $\mathbf{\Upsilon} = (\mathbf{\Upsilon}_n)_{n=1}^N$ は、低ランク近似 $\mathbf{T}\mathbf{V}$ との間に等式制約が与えられた変数であり、 $1 \leq n \leq N$ に対し $\text{rank}(\mathbf{\Upsilon}_n) \leq R$ が成り立つ。式 (4) に対応する拡張ラグランジュ関数は、

$$\begin{aligned} L_\rho(\mathbf{w}, \tilde{\mathbf{w}}, \mathbf{Y}, \mathbf{\Upsilon}, \mathbf{T}, \mathbf{V}, \tilde{\mathbf{T}}, \tilde{\mathbf{V}}, \mathbf{u}_{\tilde{\mathbf{w}}}, \mathbf{u}_{\mathbf{Y}}, \mathbf{u}_{\mathbf{\Upsilon}}, \mathbf{u}_{\mathbf{T}}, \mathbf{u}_{\mathbf{V}}) \\ = \mathcal{I}(\tilde{\mathbf{w}}) + \langle \mathbf{u}_{\tilde{\mathbf{w}}}, \mathbf{w} - \tilde{\mathbf{w}} \rangle + \frac{\rho}{2} \|\mathbf{w} - \tilde{\mathbf{w}}\|_2^2 \\ + \lambda \mathcal{D}(\mathbf{Y}, \mathbf{\Upsilon}) + \langle \mathbf{u}_{\mathbf{Y}}, \mathbf{X}\mathbf{w} - \mathbf{Y} \rangle + \frac{\rho}{2} \|\mathbf{X}\mathbf{w} - \mathbf{Y}\|_2^2 \\ + \sum_{n=1}^N \langle \mathbf{u}_{\mathbf{\Upsilon}_n}, \mathbf{\Upsilon}_n - \mathbf{T}_n \mathbf{V}_n \rangle + \sum_{n=1}^N \left(\frac{\rho}{2} \|\mathbf{\Upsilon}_n - \mathbf{T}_n \mathbf{V}_n\|_2^2 \right) \\ + \iota_+(\tilde{\mathbf{T}}) + \langle \mathbf{u}_{\mathbf{T}}, \mathbf{T} - \tilde{\mathbf{T}} \rangle + \frac{\rho}{2} \|\mathbf{T} - \tilde{\mathbf{T}}\|_2^2 \\ + \iota_+(\tilde{\mathbf{V}}) + \langle \mathbf{u}_{\mathbf{V}}, \mathbf{V} - \tilde{\mathbf{V}} \rangle + \frac{\rho}{2} \|\mathbf{V} - \tilde{\mathbf{V}}\|_2^2 \end{aligned} \quad (5)$$

Algorithm 2 Proposed Method

Input: $\mathbf{X}, \mathbf{w}^{[1]}, \tilde{\mathbf{w}}^{[1]}, \mathbf{Y}^{[1]}, \mathbf{\Upsilon}^{[1]}, \mathbf{T}_n^{[1]}, \mathbf{V}_n^{[1]}, \tilde{\mathbf{T}}_n^{[1]}, \tilde{\mathbf{V}}_n^{[1]}$,
 $\mathbf{u}_{\tilde{\mathbf{w}}}^{[1]}, \mathbf{u}_{\mathbf{Y}}^{[1]}, \mathbf{u}_{\mathbf{\Upsilon}}^{[1]}, \mathbf{u}_{\mathbf{T}}^{[1]}, \mathbf{u}_{\mathbf{V}}^{[1]}, \rho, \lambda, \alpha$

Output: $\mathbf{w}^{[K+1]}$

```

1: for  $k = 1, \dots, K$  do
2:    $\mathbf{w}^{[k+1]} = (\mathbf{X}^H \mathbf{X} + \mathbf{I})^{-1} ((\tilde{\mathbf{w}}^{[k]} - \mathbf{u}_{\tilde{\mathbf{w}}}^{[k]}) + \mathbf{X}^H (\mathbf{Y}^{[k]} - \mathbf{u}_{\mathbf{Y}}^{[k]}))$ 
3:    $\xi_{\tilde{\mathbf{w}}} = \alpha \mathbf{w}^{[k+1]} + (1 - \alpha) \tilde{\mathbf{w}}^{[k]}$ 
4:    $\tilde{\mathbf{w}}^{[k+1]} = \text{prox}_{\frac{\rho}{2}}[\xi_{\tilde{\mathbf{w}}} + \mathbf{u}_{\tilde{\mathbf{w}}}^{[k]}]$ 
5:    $\mathbf{u}_{\tilde{\mathbf{w}}}^{[k+1]} = \mathbf{u}_{\tilde{\mathbf{w}}}^{[k]} + (\xi_{\tilde{\mathbf{w}}} - \tilde{\mathbf{w}}^{[k+1]})$ 
6:   for  $n = 1, \dots, N$  do
7:     Update  $\mathbf{T}_n^{[k]}, \mathbf{V}_n^{[k]}, \tilde{\mathbf{T}}_n^{[k]}, \tilde{\mathbf{V}}_n^{[k]}, \mathbf{u}_{\mathbf{T}_n}^{[k]}, \mathbf{u}_{\mathbf{V}_n}^{[k]}$ 
       by Channel-wise Update in Alg. 3
8:   end for
9:    $\xi_{\mathbf{Y}} = \alpha (\mathbf{X}\mathbf{w}^{[k+1]}) + (1 - \alpha) \mathbf{Y}^{[k]}$ 
10:   $\xi_{\mathbf{\Upsilon}} = (1 - \alpha) \mathbf{\Upsilon}^{[k]} + \alpha \mathbf{T}^{[k+1]} \mathbf{V}^{[k+1]}$ 
11:   $[\mathbf{Y}^{[k+1]}, \mathbf{\Upsilon}^{[k+1]}] = \text{prox}_{\frac{\lambda}{\rho} \mathcal{D}}[\xi_{\mathbf{Y}} + \mathbf{u}_{\mathbf{Y}}^{[k]}, \xi_{\mathbf{\Upsilon}} - \mathbf{u}_{\mathbf{\Upsilon}}^{[k]}]$ 
12:   $\mathbf{u}_{\mathbf{Y}}^{[k+1]} = \mathbf{u}_{\mathbf{Y}}^{[k]} + (\xi_{\mathbf{Y}} - \mathbf{Y}^{[k+1]})$ 
13:   $\mathbf{u}_{\mathbf{\Upsilon}}^{[k+1]} = \mathbf{u}_{\mathbf{\Upsilon}}^{[k]} + (\mathbf{\Upsilon}^{[k+1]} - \xi_{\mathbf{\Upsilon}})$ 
14: end for

```

Algorithm 3 Channel-wise Update

Input: $\mathbf{\Upsilon}, \mathbf{T}, \mathbf{V}, \tilde{\mathbf{T}}, \tilde{\mathbf{V}}, \mathbf{u}_{\mathbf{\Upsilon}}, \mathbf{u}_{\mathbf{T}}, \mathbf{u}_{\mathbf{V}}, \alpha$

Output: $\mathbf{T}, \mathbf{V}, \tilde{\mathbf{T}}, \tilde{\mathbf{V}}, \mathbf{u}_{\mathbf{T}}, \mathbf{u}_{\mathbf{V}}$

```

1:  $\mathbf{T}_{\text{old}} = \mathbf{T}$ 
2:  $\mathbf{V}_{\text{old}} = \mathbf{V}$ 
3:  $\mathbf{T} = (\mathbf{V}\mathbf{V}^T + \mathbf{I})^{-1} (\mathbf{V}\mathbf{\Upsilon}^T + \tilde{\mathbf{T}}^T + (\mathbf{V}\mathbf{u}_{\mathbf{\Upsilon}}^T - \mathbf{u}_{\mathbf{T}}^T))^T$ 
4:  $\mathbf{V} = (\mathbf{T}^T \mathbf{T} + \mathbf{I})^{-1} (\mathbf{T}^T \mathbf{\Upsilon} + \tilde{\mathbf{V}} + (\mathbf{T}^T \mathbf{u}_{\mathbf{\Upsilon}} - \mathbf{u}_{\mathbf{V}}))$ 
5:  $\tilde{\mathbf{T}} = \max\{\mathbf{T} + \mathbf{u}_{\mathbf{T}}, 0\}$ 
6:  $\tilde{\mathbf{V}} = \max\{\mathbf{V} + \mathbf{u}_{\mathbf{V}}, 0\}$ 
7:  $\xi_{\mathbf{T}} = (1 - \alpha) \mathbf{T}_{\text{old}} + \alpha \tilde{\mathbf{T}}$ 
8:  $\xi_{\mathbf{V}} = (1 - \alpha) \mathbf{V}_{\text{old}} + \alpha \tilde{\mathbf{V}}$ 
9:  $\mathbf{u}_{\mathbf{T}} = \mathbf{u}_{\mathbf{T}} + (\mathbf{T} - \xi_{\mathbf{T}})$ 
10:  $\mathbf{u}_{\mathbf{V}} = \mathbf{u}_{\mathbf{V}} + (\mathbf{V} - \xi_{\mathbf{V}})$ 

```

となる。ただし、 $\mathbf{u}_{\tilde{\mathbf{w}}}, \mathbf{u}_{\mathbf{Y}}, \mathbf{u}_{\mathbf{\Upsilon}}, \mathbf{u}_{\mathbf{T}}, \mathbf{u}_{\mathbf{V}}$ は双対変数である。ADMM では、8 つの主変数のそれぞれに関して式 (5) を交互に最小化し、続いて 5 つの双対変数のそれぞれにおいて勾配上昇を行う。これらの操作を Alg. 2, 3 に記す。ただし、 $\xi_{\tilde{\mathbf{w}}}, \xi_{\mathbf{Y}}, \xi_{\mathbf{\Upsilon}}, \xi_{\mathbf{T}}, \xi_{\mathbf{V}}$ の更新式はアルゴリズムの安定性に関わる緩和操作であり、 $\alpha \in (0, 2)$, $\rho > 0$, $\lambda > 0$ である [7]。また、 $\max\{\cdot, 0\}$ は要素ごとの非負射影である。

提案手法は、正則化項や制約の追加が容易である。追加された正則化項に関しては、補助変数を導入して変数分離を行い、得られた問題に対応する拡張ラグランジュ関数を再度導出する。その後、これを各変数について交互最適化することでアルゴリズムが得られる。追加された制約に対しては、指示関数を用いて再定式化することで、正則化項の追加と同様に対応できる。また、数値安定性に関しても、NMF による低ランク近似や分離行列の更新における不安定な逆行列演算¹や除算は回避されている。

¹ $\mathbf{X}^H \mathbf{X}, \mathbf{V}\mathbf{V}^T, \mathbf{T}^T \mathbf{T}$ はいずれも半正定値対称行列であり、単位行列が足されることで正定値対称行列となるため逆行列演算の数値不安定性が解消される。また、 $\mathbf{X}^H \mathbf{X} \in \mathbb{C}^{NMF \times NMF}$ は M 次正方行列が並んだブロック対角行列で、 $\mathbf{V}\mathbf{V}^T, \mathbf{T}^T \mathbf{T}$ はいずれも R 次正方行列と比較的小さな正方行列であり、容易に計算できる。

3.2 Perspective 関数による ILRMA 型 BSS

式 (4) 中の関数 \mathcal{D} には、分離音 \mathbf{Y} の振幅スペクトログラムと低ランク近似 \mathbf{Y} の間に距離やそれに類似した尺度を与える関数を用いる。ただし、従来の ILRMA で用いられる局所ガウス分布の負対数尤度関数に対応する関数は、2変数を同時最適化する近接作用素を解析的に求めることができない。そこで本稿では、 \mathbf{Y} と \mathbf{Y} の構造を関連付ける [8] 関数として、以下で定義される凸関数 Φ を用いる。

$$\Phi(\mathbf{Y}, \mathbf{Y}) = \sum_{t=1}^T \sum_{f=1}^F \sum_{n=1}^N \phi(y_n[t, f], v_n[f, t]) \quad (6)$$

ただし、 ϕ は $|\cdot|^2/2 + 1/2$ の perspective 関数であり、

$$\phi(y, v) = \begin{cases} \frac{|y|^2}{2v} + \frac{v}{2} & (v > 0) \\ 0 & (y = 0 \text{ and } v = 0) \\ +\infty & (\text{otherwise}) \end{cases} \quad (7)$$

で定義される [4]。関数 Φ は近接作用素が解析的に求められるため [4, 9]、更新式の導出が容易である。また、局所ガウス分布に基づく関数は非凸であるため、凸関数 Φ で置き換えることで、初期値依存性の低減が期待できる。関数 Φ の近接作用素 $\text{prox}_{\tau\Phi}$ は、要素ごとに計算可能であり、以下の式で書ける。

$$\text{prox}_{\tau\Phi}[\mathbf{Y}, \mathbf{Y}] = (((\text{prox}_{\tau\phi}[y_n[t, f], v_n[f, t]])_{t=1}^T)_{f=1}^F)_{n=1}^N \quad (8)$$

また、近接作用素 $\text{prox}_{\tau\phi}$ は解析的に得られ、

$$\text{prox}_{\tau\phi}[y, v] = \begin{cases} (0, 0) & (2\tau v + |y|^2 \leq \tau^2) \\ (0, v - \frac{\tau}{2}) & (y = 0 \text{ and } 2v > \tau) \\ \left(y - \tau s \frac{y}{|y|}, v + \tau \frac{s^2 - 1}{2}\right) & (\text{otherwise}) \end{cases} \quad (9)$$

と書ける [4]。ただし、 $s > 0$ は、

$$s = \begin{cases} \sqrt[3]{-\frac{q}{2} + \sqrt{-r}} + \sqrt[3]{-\frac{q}{2} - \sqrt{-r}} & (r < 0) \\ 2\sqrt[3]{-\frac{q}{2}} & (r = 0) \\ 2\sqrt[3]{\sqrt{\frac{q^2}{4} + r} \cos\left(\frac{\arctan(-2\frac{\sqrt{r}}{q})}{3}\right)} & (r > 0) \end{cases} \quad (10)$$

と書ける [9]。ここで、 $p = (2/\tau)v + 1$, $q = -(2/\tau)|y|$, $r = -q^2/4 - p^3/27$ であり、 $\sqrt[3]{\cdot}$ は実立方根である。Alg. 3 中の近接作用素 $\text{prox}_{\frac{\lambda}{\rho}\mathcal{D}}$ を、関数 Φ の近接作用素 $\text{prox}_{\frac{\lambda}{\rho}\Phi}$ で置き換えることで、perspective 関数を用いた ILRMA 型 BSS アルゴリズムが得られる。

4 実験

提案手法 (ADMM-ILRMA) の有効性を確認し、提案手法における適切なパラメータの値を検討するため、AuxILRMA (IP[10], ISS[11]) との比較実験を行った。

4.1 実験条件

データセットは SiSEC 2011 の dev1 を用いた。残響時間は 130 ms, 250 ms, マイク間距離は 5 cm, 1 m とし、女声男声それぞれ 3 組を用いて 2 チャネルの 2 話者混合音 (24 組) を作成した。サンプリング周波数は 16 kHz, STFT の窓関数は 2048 サンプルの Hann 窓, シフト長は 1024 サンプルとした。反復回数はすべての手法で 1000 回とし、評価指標には ΔSDR を用いた。基底行列 \mathbf{T} およびアクティベーション行列 \mathbf{V} の初期値は 0 から 1 の一様分布乱数とし、乱数シードを変えて各手法をそれぞれ 10 回試行した。基底ベクトルのランク数 R は 2 とした。提案手法におけるパラメータのうち、ステップサイズに関するパラメータ ρ を 0.05, \mathcal{D} の重み λ を 1, 0.1, 0.01, 0.001, アルゴリズムの緩和に関するパラメータ α を 1.0, 1.1, 1.2, 1.3, 1.4, 1.5 とした。

4.2 パラメータ λ に関する検討

まず、 \mathcal{D} の重み λ に関する検討を行う。最終的な ΔSDR の値を図-1 に示す。提案手法では、 $\lambda = 0.01$ で従来手法に近い性能が得られた。このとき、従来手法と比較して外れ値が少なく、分離に失敗するケースを回避することが分かった。特に $\alpha = 1.1$ の時、最も高い中央値が得られた。全体的な傾向として、提案手法は従来手法と比較して平均値と中央値の差が小さかった。外れ値が減少した理由として、凸な正則化項を用いたことで初期値依存性が低減されたことや、提案フレームワークが数値不安定性の要因となる演算を含まないことが考えられる。

4.3 パラメータ α に関する検討

次に、従来手法に近い性能が得られた $\lambda = 0.01$ の場合において、アルゴリズムの緩和に関する最適化パラメータ α の値を検討する。 ΔSDR の平均値の推移を図-2 に示す。提案手法では、4.2 節と同様に $\alpha = 1.1$ とした時に最も高い平均値が得られ、400 反復程度で従来手法の平均値を上回った。

4.4 NMF による低ランク近似の比較

最後に、提案手法において最も良い分離性能が得られた $\lambda = 0.01$, $\alpha = 1.1$ の場合について、その NMF による低ランク近似を従来手法と比較した。図-3 に示すように、従来手法と同様に、提案手法は分離音の振幅の低ランク近似を得ることが確認できた。

5 むすび

本稿では、ADMM を用いた ILRMA 型 BSS フレームワークを提案し、正則化項として perspective 関数を用いた ILRMA を実現した。実験の結果、提案手法は適切なパラメータの下で従来手法と同程度の分

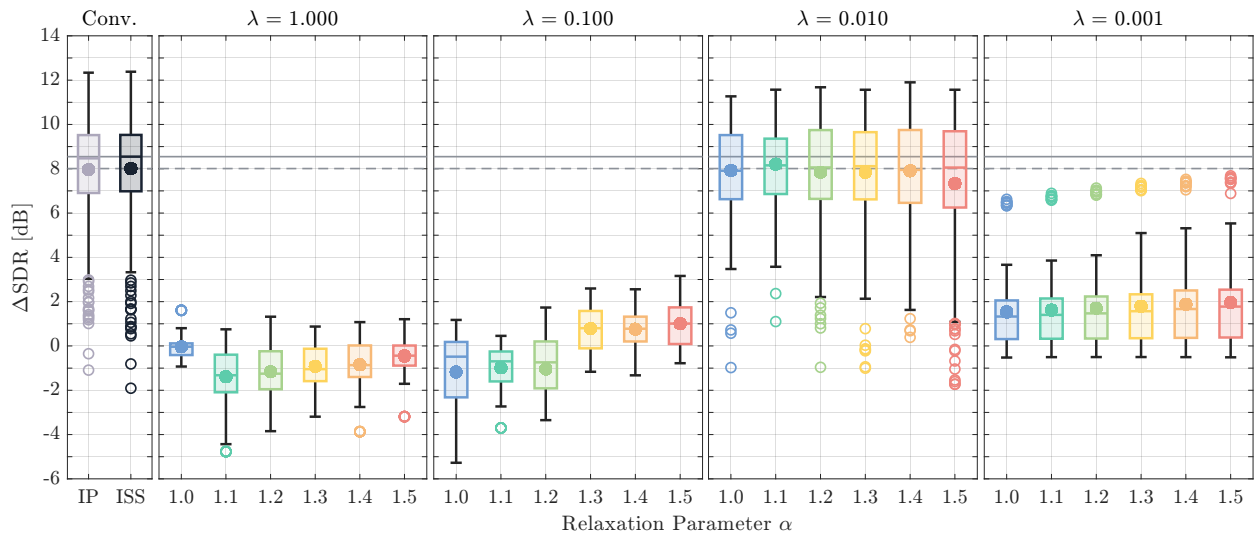


図-1 1000 反復後の分離性能 (ΔSDR). 1 列目は従来手法の結果である. それ以降は提案手法の結果であり, それぞれの λ について各 α の性能を示している. 各手法における ΔSDR の平均値を塗りつぶした丸としてプロットしている. 図を横断する実線および点線は AuxILRMA-ISS の中央値および平均値をそれぞれ示している.

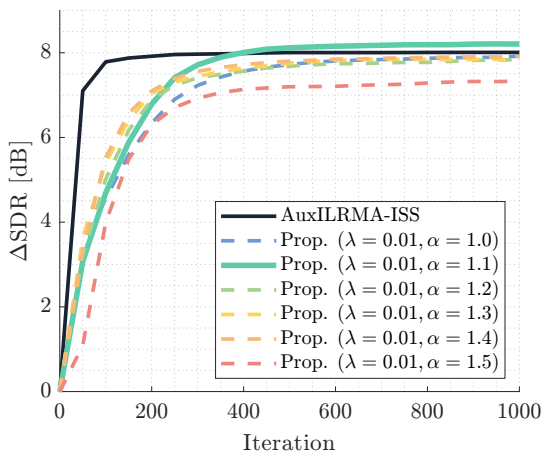


図-2 反復に伴う ΔSDR の平均値の推移. ΔSDR の平均値を 50 反復ごとに示している. 1000 反復後の ΔSDR の平均値は, 従来手法では 8.01 dB, 提案手法 ($\lambda = 0.010, \alpha = 1.1$) では 8.21 dB であった.

離性能を実現し, 従来手法と比較して分離に失敗するケースを減らせることが分かった. その要因として, perspective 関数の凸性による初期値依存性の低減や, 数値的に不安定な演算の回避が考えられた. 今後の展望として, 初期値に関する更なる検証や, 提案手法の柔軟性を活かした正則化項や制約の追加, 提案フレームワークを用いた局所ガウス分布に基づく ILRMA の実現などが挙げられる.

参考文献

[1] T. Kim, T. Eltoft and T. W. Lee, “Independent Vector Analysis: An Extension of ICA to Multivariate Components,” *Indep. Compon. Anal. Blind Signal Sep.*, Vol. 3889, pp. 165–172 (2006).
 [2] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, “Determined Blind Source Separation Unifying Independent Vector Analysis and Nonnegative Matrix Factorization,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **24**, 1626–1641 (2016).
 [3] 渡會博子, 山田宏樹, 矢田部浩平, “ADMM アルゴリズムを用いた優決定ブラインド音源分離,” 音講論集, pp. 145–146

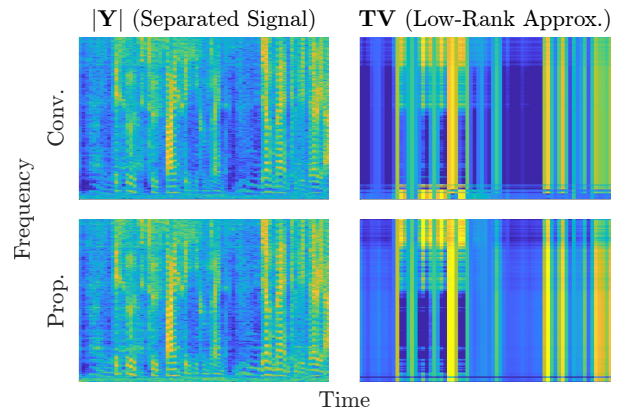


図-3 分離音の振幅 $|\mathbf{Y}|$ (1 列目) と低ランク近似行列 \mathbf{TV} (2 列目) の比較. 1 行目に従来手法 (AuxILRMA-ISS) の結果, 2 行目に提案手法の結果を示している. 視認性のために, 一部を拡大表示し, 色の範囲を調整している.

(2023.9).
 [4] P. L. Combettes and C. L. Müller, “Perspective Maximum Likelihood-type Estimation via Proximal Decomposition,” *Electron. J. Stat.*, **14**, 207–238 (2020).
 [5] K. Yatabe and D. Kitamura, “Determined BSS Based on Time-Frequency Masking and Its Application to Harmonic Vector Analysis,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **29**, 1609–1625 (2021).
 [6] D. L. Sun and C. Févotte, “Alternating Direction Method of Multipliers for Non-negative Matrix Factorization with the Beta-divergence,” *IEEE ICASSP*, pp. 6201–6205 (2014).
 [7] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers* (now, 2011).
 [8] K. Arai, K. Yamada and K. Yatabe, “Versatile Time-Frequency Representations Realized by Convex Penalty on Magnitude Spectrogram,” *IEEE Signal Process. Lett.*, **30**, 1082–1086 (2023).
 [9] H. Kuroda and K. Daichi, “Block-Sparse Recovery With Optimal Block Partition,” *IEEE Tran. Signal Process.*, **70**, 1506–1520 (2022).
 [10] N. Ono, “Stable and Fast Update Rules for Independent Vector Analysis Based on Auxiliary Function Technique,” *IEEE WASPAA*, pp. 189–192 (2011).
 [11] S. Robin and N. Ono, “Fast and Stable Blind Source Separation with Rank-1 Updates,” *IEEE ICASSP*, pp. 236–240 (2020).